

## PENERAPAN PENDEKATAN MACHINE LEARNING PADA PENGEMBANGAN BASIS DATA HERBAL SEBAGAI SUMBER INFORMASI KANDIDAT OBAT KANKER

### THE APPLICATION OF MACHINE LEARNING APPROACH TOWARDS HERBAL DATABASE DEVELOPMENT AS INFORMATION SOURCE FOR CANCER DRUG CANDIDATE

Arli Aditya Parikesit<sup>\*</sup>), Rizky Nurdiansyah, dan David Agustriawan

Department of Bioinformatics, School of Life Sciences, Indonesia International Institute for Life Sciences,  
Jl. Pulomas Barat Kav 88 Jakarta 13210  
Email: [arli.parikesit@i3l.ac.id](mailto:arli.parikesit@i3l.ac.id)

Makalah: Diterima 11 Januari 2019; Diperbaiki 23 Juli 2019; Disetujui 3 Agustus 2019

#### ABSTRACT

Cancer is still an epidemiological disease in Indonesia. Drug development against cancer still relies to pharmacological laboratories and natural chemicals, which could have side effects. Cancer drug development has entered the stage of molecular biology, where the interaction of ligand chemical structure with receptor protein can be studied with high accuracy. Various chemical compounds, ranging from synthetic, semi-synthetic, to natural materials, developed for the purpose to fight one of the most dangerous diseases. In the context of the development of herbal-based drugs, there has been found heaps of natural compounds, curated and annotated, in various databases belonging to China, Taiwan, Indonesia, Japan, and several other countries. However, problems arise when choosing the best bioactive compounds to develop against cancer. Complexity arises because the metabolic pathway of cancer is very diverse, depending on the type and phase of cancer. Therefore, in this systematic review, we developed a machine learning approach to screen for these bioactive compounds, then took the best candidates for molecular simulation operations that would be tested for validity in wet experiments. Thus, the automation of the candidate drug development process for cancer could be achieved with great significance. It is known that the most effective and efficient machine learning method was Naïve Bayes, but the best in processing large amounts of compound data was classifier SVM. The future of complex bioactive compounds data could be secured by employing deep learning method.

Keywords: machine learning, drug development, natural material compounds, metabolic pathways, cancer

#### ABSTRAK

Penyakit kanker merupakan masalah epidemiologis di tanah air. Pengembangan obat modern sejauh ini masih bergantung pada laboratorium farmakologi dan kimia bahan alam yang dapat memiliki efek samping. Namun, pengembangan obat untuk penyakit kanker sudah memasuki tahap biologi molekuler, dimana interaksi struktur kimia ligan dengan protein reseptor dapat dikaji dengan ketelitian tinggi. Berbagai tipe senyawa kimia, mulai dari sintetik, semi sintetik, hingga bahan alam, dikembangkan untuk keperluan melawan penyakit yang dianggap paling berbahaya tersebut. Dalam konteks pengembangan obat berbasis herbal, telah ditemukan senyawa bahan alam dalam jumlah banyak. Data tersebut dikurasi dan dianotasi oleh basis data milik China, Taiwan, Indonesia, Jepang, maupun beberapa negara lainnya. Hanya saja, permasalahan timbul ketika memilih senyawa bioaktif terbaik untuk dikembangkan untuk melawan penyakit kanker. Kompleksitas timbul karena jalur metabolik penyakit kanker sangat beragam, tergantung pada tipe dan fasenya. Oleh karena itu, dalam telaah sistematis ini, disajikan telaah pendekatan *machine learning* untuk melakukan penapisan terhadap pustaka senyawa bioaktif tersebut, kemudian proses seleksi kandidat yang terbaik untuk operasi simulasi molekuler, dan selanjutnya teruji validitasnya pada *wet experiment*. Sehingga proses otomatisasi pengembangan kandidat obat bagi penyakit kanker dapat dicapai dengan sangat signifikan. Diketahui bahwa metode *machine learning* paling efektif dan efisien adalah *Naïve Bayes*. Namun yang terbaik dalam mengolah data senyawa dalam jumlah besar adalah *SVM classifier*. Kedepannya, metode *deep learning* sangat menjanjikan untuk komputasi data senyawa bioaktif yang kompleks.

Kata kunci: machine learning, pengembangan obat, senyawa bahan alam, jalur metabolik, penyakit kanker

#### PENDAHULUAN

Kanker merupakan penyakit yang mematikan yang disebabkan oleh proliferasi abnormal sel-sel tubuh. Sel-sel abnormal tersebut dapat merusak fungsi organ dan menyebabkan kematian. Insiden tahunan kanker diseluruh dunia

kurang lebih sekitar 30,000 dimana jenis kanker yang paling mematikan adalah kanker paru-paru dan kanker payudara (Lim, 2002). Dengan meningkatnya prevalansi kanker, pengobatan komplementer dan alternatif sama-sama dibutuhkan sebagai obat kanker (Chrystal *et al.*, 2003). Obat herbal dan tanaman seperti buah-buahan, sayur-sayuran, dan spesies

<sup>\*</sup>Penulis Korespondensi

tumbuhan lainnya memiliki banyak khasiat pada kesehatan (Lampe, 1999; Surh, 2003). Saat ini operasi, radioterapi dan kemoterapi merupakan pilihan utama untuk pengobatan kanker, sedangkan bahan senyawa alam masih dianggap sebagai tambahan pengobatan (Golbeck *et al.*, 2011). Uniknya, 60 sampai 75% dari obat anti kanker merupakan produk dari senyawa alam. Obat yang berasal dari senyawa alam diklaim lebih manjur untuk pasien kanker dibandingkan dengan obat sintetis. Untuk saat ini, diketahui bahwa 10,000 dari 50,000 spesies tanaman yang mengandung khasiat obat, tersebar di berbagai ekosistem, namun hanya sebagian dari tanaman tersebut telah dianalisa dan diinvestigasi untuk potensi obat terapi.

Di Malaysia, studi tanaman herbal sebagai sumber potensial dari kanker terapi sedang dikembangkan. Di antara tanaman-tanaman yang sedang diinvestigasi untuk agen terapi potensial adalah Keladi tikus (*Typhonium flagelliforme*) untuk pengobatan leukemia (Mohan *et al.*, 2010(a); Mohan *et al.*, 2010(b)), *chalcone* dari Temu Kunci (*Boesenbergia rotunda*) untuk kanker paru-paru (Isa *et al.*, 2012), dan Tapak Liman (*Elephantopus scaber*) untuk kanker payudara (Ho *et al.*, 2011). Beberapa zat aktif lainnya yang diinvestigasi adalah *girinimbine* dari akar dari Salam Koja (*Muraya koenigii*) untuk pengobatan kanker hati (Syam *et al.*, 2011), *dentatin* dari daun Sicerek (*Clausena excavate burm*) untuk kanker prostat (Arbab *et al.*, 2013), minyak biji dari Kenaf (*Hibiscus cannabinus*) dan *phenylbutenoids* dari Bangle (*Zingiber cassumunar roxb*) untuk leukemia (Foo *et al.*, 2011; Anassamy *et al.*, 2013). Indonesia setidaknya memiliki 9,600 dari sekitar 30,000 spesies yang mempunyai aktivitas farmakologis (Depkes RI, 2007). Penelitian sebelumnya telah membuat basis data senyawa herbal yang berasal dari Indonesia, Basis data tersebut menyediakan informasi dari struktur tiga dimensi dari senyawa herbal yang dapat diakses di tautan berikut: <http://herbaldb.farmasi.ui.ac.id>. Penelitian sebelumnya juga meneliti senyawa herbal yang terdapat di Vietnam dan mengumpulkan struktur kimia dari senyawa tersebut dari situs *Pubchem* dan *Chempider database* (<http://pubchem.ncbi.nlm.nih.gov> dan <http://www.chemspider.com/>) (Ngo dan Li, 2013). Dalam mendesain suatu obat, tanaman-tanaman herbal ini menjadi bahan dasar untuk dipelajari lebih jauh. Penyaringan dari kegiatan farmakologi dari zat aktif tanaman obat memerlukan biaya yang besar, sumber daya manusia yang handal, dan waktu yang lama jika dilakukan tanpa arah yang jelas (Jayaram, 2011). Metode *in silico* atau ilmu bioinformatika dapat membuat simulasi dan perhitungan didalam desain obat. Metode yang digunakan adalah *Computer Assisted Drug Design* (CADD) yang mengkaji interaksi antara kandidat obat dan target secara komputasi (Hawkins dan Skillman, 2006). *Molecular docking* adalah metode umum yang

digunakan. *Molecular docking* digunakan untuk memprediksi kompleks intermolekuler antara molekul obat dan target protein. Data yang dibutuhkan terdiri dari informasi ligan, atau obat yang akan dianalisis, dan reseptor atau protein target. Informasi yang dibutuhkan pada proses ini merupakan struktur tiga dimensi dari ligan dan reseptor (Abraham, 2003). Kemudian, ilmu bioinformatika juga menggunakan pendekatan *machine learning* untuk mempelajari kanker dan virus, sehingga hal ini dapat menjadi modal informasi yang baik untuk pengembangan obat berbasis herbal (Parikesit, 2018; Parikesit *et al.*, 2018).

Telaah sistematis ini akan membahas studi kasus yang memanfaatkan senyawa herbal untuk pengobatan kanker berbasis metode *machine learning*. Salah satu contoh yang akan didalami adalah *Lysine-specific demethylase 1* (LSD1) yang telah diketahui mengalami ekspresi yang lebih tinggi di banyak jenis sel kanker, diantaranya pada kanker payudara (Lim *et al.*, 2010), neuroblastoma (Schulte *et al.*, 2009), dan kanker prostat (Wissman *et al.*, 2007). Dengan demikian, penghambatan terhadap LSD1 oleh molekul lainnya adalah strategi yang efektif untuk menurunkan ekspresi gen yang terlibat pada proses diferensiasi, migrasi, dan invasi dari sel kanker atau untuk menghambat jalur proliferasi di berbagai kanker (Mohammad *et al.*, 2013; Schmitt *et al.*, 2013). *Xanthones* adalah senyawa polifenol alami dengan beragam aktivitas biologi, biokimia, dan farmakologi yang ditemukan di berbagai macam tanaman (Negi *et al.*, 2013).

Salah satu jenis dari *xanthones*,  *$\alpha$ -mangostin* adalah konstituen utama yang diisolasi dari buah manggis (*Garcinia mangostana*). Khasiat biologis yang signifikan menjadikannya sebagai kandidat bahan untuk pengembangan obat yang baru, khususnya untuk obat anti-kanker (Ibrahim *et al.*, 2015). Dengan demikian, telaah ini menjadikan  *$\alpha$ -mangostin* sebagai inhibitor terhadap LSD1. Kajian bioinformatika, diperkuat dengan kajian laboratorium basah, telah dilakukan dalam mengkaji beberapa senyawa bahan alam dan turunannya sebagai inhibitor LSD1. Diantaranya adalah sintesis turunan *Xanthine* (Ma *et al.*, 2019); kajian simulasi molekuler terhadap turunan senyawa *6-aril-5-siano-pirimidin* (Ding *et al.*, 2017); dan simulasi molekuler terhadap relasi ROR  $\alpha$  terhadap LSD1 (Kim *et al.*, 2017). Keberhasilan penemuan ligan untuk inhibitor LSD1 dengan metode komputasi, terutama yang berbasis bahan alam dan turunannya, merupakan motivasi utama dalam kajian telaah sistematis ini. Sehingga diharapkan dapat memberikan narasi yang jelas mengenai detail teknis dari metode *machine learning* yang digunakan dalam simulasi tersebut dalam contoh interaksi reseptor-ligan dalam kasus lainnya secara lebih umum.

### Navigasi Koleksi Struktur 3D dari Bahan Alam

Langkah pertama untuk mendesain obat molekuler dari bahan alam adalah pengumpulan informasi terhadap bahan alam yang dapat digunakan sebagai obat. Informasi tersebut dapat ditemukan pada literatur, jurnal, buku dan situs-situs. Seperti yang telah dijelaskan dalam telaah Parikesit *et al.* (2018), langkah ini dimulai dengan mengumpulkan informasi mengenai bahan kimia yang ditemukan di tanaman obat beserta struktur 2D dari berbagai literatur dan situs *Pubchem*. Kemudian, *software* PyMOL digunakan untuk menghasilkan struktur 3-dimensi dan menyimpannya ke dalam basis data. Informasinya tersedia di <http://herbaldb.farmasi.ui.ac.id>. Lebih dari 1412 struktur 3-dimensi dari bahan kimia dari bahan alam di Indonesia yang disimpan di situs <http://herbaldb.farmasi.ui.ac.id> yang dapat diakses oleh publik dengan persetujuan admin untuk penelitian (Yanuar *et al.*, 2011).

### Analisis *in silico* atau Bioinformatika

Setelah mengetahui struktur dari bahan alam dan target protein, pengujian atau analisis *in silico* dapat diterapkan untuk melihat seberapa signifikan pengaruh bahan alam tersebut terhadap target protein. Dari penelitian yang dilakukan oleh Han *et al.* (2018), langkah-langkah yang dilakukan untuk mengetahui efek dari  *$\alpha$ -mangostin* terhadap LSD1 dalam aktivitas pengobatan kanker adalah analisis penambatan molekul dan pendekatan *machine learning*.

### Penambatan Molekul

Analisis penambatan molekul dilakukan menggunakan *software* MOE 2009. Dalam kasus LSD1, Struktur dari LSD1 (PDB ID: 2V1D) dengan FAD (flavin adenine dinucleotide) bebas, sebuah peptida yang mirip dengan histone H3k4 methyltransferase sebagai substratnya, dan CoREST (Protein REST corepressor 1) sebagai korepresor dipilih sebagai protein untuk analisis *docking*. Posisi grup dihidroksil dari rangka *xanthone* membentuk ikatan hidrogen dengan rantai samping dari Arg316 dan Thr810. Cincin benzen membentuk interaksi  $\pi$ - $\pi$  dengan Tyr761, Glu801, Ala331, Val811, dan Arg316. Dua kelompok isopentene diprediksi berikatan dengan interaksi hidrofobik dengan Leu659, Trp751, dan Ala814. Dari analisis penambatan molekul diatas, rangka *xanthone* ditemukan secara signifikan berinteraksi dengan LSD1 dan kedua kelompok isopentene dapat berfungsi sebagai kelompok fungsional aktif untuk mempengaruhi bioaktivitas dari LSD1 (Han *et al.*, 2018). Tidak hanya pada LSD1, metode ini juga dapat digunakan untuk mengkaji penambatan molekul ligan-protein lainnya.

### Pendekatan *Machine Learning*

Penggunaan pendekatan *machine learning* telah digunakan pada berbagai aspek penelitian yang

berhubungan dengan pengembangan obat. *Machine learning* adalah pendekatan prediktif yang mampu memprediksi luaran suatu sistem berbasis pada pola yang sudah ada. Runutan dalam metode *machine learning* dijabarkan sebagai berikut: Kumpulan besar data senyawa dari basis data bahan alam yang tersedia untuk umum diambil (proses *data mining*) untuk mengidentifikasi senyawa terapi potensial untuk kanker. Jutaan senyawa dapat disaring untuk mengkaji aktivitas anti-kanker oleh *custom build classifier SVM*. Target molekuler senyawa anti-kanker yang diprediksi didapatkan dari sumber yang dapat diandalkan, seperti penelitian *bioassay* eksperimental yang terkait dengan senyawa tersebut dan dari database interaksi protein-ligan. Basis data senyawa yang dapat digunakan sebagai terapi dan daftar senyawa anti-kanker bahan alam yang berasal dari studi literatur digunakan untuk membangun model regresi kuadrat parsial terkecil. Model regresi yang dibangun digunakan untuk estimasi bobot spesifik kanker berdasarkan target molekuler. Bobot ini digunakan untuk menghitung skor untuk menyaring senyawa anti-kanker yang diprediksi memiliki potensi dalam mengobati kanker (Bundela *et al.*, 2015). Pada telaah sistematis ini, beberapa penelitian yang sesuai telah ditemukan dalam kurun waktu 10 tahun terakhir. Beberapa dilaporkan dengan penggunaan data visual dan *natural language processing* dengan menerapkan konsep *data mining* untuk melakukan identifikasi tanaman obat maupun kandungan zat aktif yang dimiliki.

### Penggunaan *Machine Learning* Dalam Identifikasi Tanaman Obat

Proses identifikasi dan klasifikasi spesies yang memiliki khasiat obat merupakan langkah awal dari proses penemuan obat. Proses identifikasi membutuhkan tenaga ahli yang memahami penanda yang dimiliki oleh spesies, baik secara morfologi maupun genetis. Sejauh ini, proses yang paling sukses dalam identifikasi adalah menggunakan pendekatan morfologi, meskipun melalui proses yang cukup panjang (Begue *et al.*, 2017). Pendekatan *machine learning* telah digunakan dalam melakukan proses identifikasi tanaman berikut ini (Tabel 1).

Dalam kurun 10 tahun, pendekatan *machine learning* telah membantu dalam proses identifikasi secara visual dan bahkan telah berhasil diterapkan untuk mendeteksi lesi, mitosis, dan status gen pada kanker payudara (Ramanto dan Parikesit, 2019). Sehingga validitas dan reliabilitasnya tak perlu diragukan lagi. Berdasarkan tabel 1, terdapat metode *machine learning* yang berbeda dan memberikan akurasi yang cukup baik dalam mengidentifikasi spesies yang digunakan dalam proses *training*. Du *et al.* (2009) menggunakan metode kNN untuk mengidentifikasi 2000 foto dari 20 spesies dan mendapatkan 92,3% akurasi. Penelitian lainnya menggunakan metode seperti PNN, ANN, dan SVM

yang memberikan akurasi melebihi 90% (Hossain dan Amin, 2010; Le *et al.*, 2014; Chaki *et al.*, 2015). Salah satu metode yang dikembangkan bahkan dibuat menjadi aplikasi untuk *smartphone* seperti MedLeaf (Prasvita dan Herdiyeni, 2013). Metode *machine learning* dapat diterapkan pada tanaman yang berasal dari manapun, seperti Indonesia (Herdiyeni dan Wahyuni, 2012; Prasvita dan Herdiyeni, 2013), Vietnam (Le *et al.*, 2014), bahkan tanaman ornamental (Arai *et al.*, 2013) selama basis data yang digunakan untuk “melatih” program memiliki informasi yang cukup. Unikanya, kebanyakan dari program yang dibuat, menggunakan data yang sama dengan data yang digunakan untuk “melatih” program *Flavia* yang berisi 33 gambar daun tanaman yang diketahui memiliki khasiat obat (Wu *et al.*, 2007). Dengan menggunakan *dataset* tersebut, peneliti lain mengembangkan model ResNet hingga 26 lapisan sehingga dapat mengidentifikasi tanaman di habitat asli dengan akurasi 91,78% dan data daun pada *Flavia* hingga 99,65% (Yu *et al.*, 2017). Selain untuk identifikasi tanaman, metode *machine learning* juga digunakan untuk analisis senyawa anti-kanker yang disajikan pada Tabel 2.

Berdasarkan Tabel 2, metode berbasis *machine learning* pada dasarnya terbagi dua, yaitu untuk data senyawa berukuran kecil/sedang (nomor 1-3), dan berukuran besar (nomor 4). Salah satu alasan mengapa *SVM classifier* merupakan metode yang sangat cocok untuk data berukuran sangat besar adalah karena metode tersebut secara ketat membangun model berdasarkan contoh dari data pelatihan. Selanjutnya, data tersebut digunakan untuk memprediksi nilai target dari contoh data uji dan hanya diberikan atribut dalam data uji (Bundela *et al.*, 2015). Sementara itu, metode SPIDER menggunakan *self organizing map* (SOM) yang berbasis algoritma kluster sehingga cenderung membutuhkan daya komputasi besar untuk set data yang kecil (Reker *et al.*, 2014). *Naive Bayes*, *Random Forest*, *j48*, *SMO* adalah algoritma yang paling populer digunakan pada kajian immunoinformatika, karena sudah sangat tervalidasi (Jamal dan Scaria, 2013). Diantara mereka, algoritma *Naive Bayes* adalah yang paling cepat (Kukreja *et al.*, 2012). Hanya saja algoritma tersebut cenderung kurang baik akurasi pada data yang sangat besar ukurannya.

Tabel 1. Pendekatan *machine learning* pada identifikasi tanaman (dimodifikasi dari Begue *et al.*, 2017).

Referensi	Fitur	Metode*	Akurasi (%)	Spesies
Du <i>et al.</i> (2009)	Bentuk	kNN	92,3	20
Backes <i>et al.</i> (2009)	Tekstur	LDA	89,6	10
Hossain dan Amin (2010)	Bentuk	PNN	91,4	30
Du <i>et al.</i> (2013)	Lekukan, lapisan	kNN	87,1	30
Amin and Khan (2013)	Lekukan	kNN	71,5	100
Herdiyeni and Wahyuni (2012)	Tekstur, warna	PNN	74,5	51
Arai <i>et al.</i> (2013)	<i>Discrete wavelets</i>	SVM	95,8	8
Hernandez-Serna and Jimenez-Segura (2014)	Bentuk, tekstur	ANN	92,9	32
Le <i>et al.</i> (2014)	<i>Kernel Descriptor</i>	SVM	98,4	32
Munisami <i>et al.</i> (2015)	Bentuk, warna	kNN	87,3	32
Chaki <i>et al.</i> (2015)	Bentuk, tekstur	NFC	97,6	31
Siravenha dan Carvalho (2015)	Bentuk	ANN	97,5	32
Carranza-Rojas dan Mata-Montero (2016)	Lekukan, tekstur	kNN	87,2	66
Begue <i>et al.</i> (2017)	Bentuk	RF	90,1	24

\*kNN: *k-nearest neighbor*; SVM: *support vector machine*; PNN: *Probabilistic neural network*; NFC: *Neuro-fuzzy classifier*; ANN: *Artificial neural network*; RF: *Random forest*.

Tabel 2. Pendekatan *machine learning* pada penapisan senyawa anti-kanker

No.	Referensi	Tipe Kanker	Biomolekul Target	Metode	Jumlah Senyawa Bahan Alam/ Turunannya yang Ditapiskan
1	(Wahi <i>et al.</i> , 2015)	Belum Ditentukan**	USP1/UAF1	Naive Bayes, Random Forest, j48, SMO	13931
2	(Nand <i>et al.</i> , 2016)	Paru-paru	EGFR	Random Forest, J48, and Bayes Net	419
3	(Schneider <i>et al.</i> , 2016)	Belum Ditentukan**	EP3	Spider Protocol	135
4	(Bundela <i>et al.</i> , 2015)	Kanker mulut	Bervariasi, diunduh dari basis data STITCH	SVM Classifier	84 Million

\*\*Pengujian senyawa tersebut menggunakan penapisan galur sel kanker (*cell line*)

## Penggunaan *Machine Learning* Meningkatkan Efisiensi Penemuan Obat

Penggunaan model prediktif dengan menggunakan komputer dapat membantu memprediksi hasil eksperimen yang bahkan belum dilaksanakan. Bila model tersebut telah dibuat, maka perlu adanya eksperimen lain untuk mengetahui akurasi dari model. Pembuatan model tersebut harus secara teliti karena hasil *false-positive* maupun *false-negative* dapat memberikan kerugian yang signifikan. (Murphy *et al.*, 2013). Beberapa pendekatan dilakukan untuk meningkatkan akurasi dari model prediktif. Salah satunya adalah algoritma *active learning* yang mengedepankan otomatisasi dalam proses “pelatihan”. Metode ini dapat memprediksi 60% adanya reaksi dari data senyawa dan protein setelah melakukan eksplorasi sebanyak 3% dari total 3,54 juta kemungkinan, jauh lebih baik dari model *greedy selection with QSAR-like model* dan *Computational Research Engine*. Uniknya, model ini akan semakin baik seiring dengan semakin banyaknya data yang diberikan (Naik *et al.*, 2013).

Model *machine learning* juga dibuat untuk mengidentifikasi adanya ikatan antara senyawa aktif dengan protein. Berbeda dengan sebelumnya, model ini menggunakan *cascaded learning method* yang terdiri atas 2 tahapan. Tahapan pertama menggunakan metode yang menyeleksi *binding activity* dari senyawa berdasarkan strukturnya. Hal ini didasarkan pada metode *structure-activity-relationship* (SAR) yang dapat digunakan untuk memprediksi ikatan dari suatu senyawa, selama diketahui strukturnya. Tahapan kedua menggunakan data SAR dari tahap pertama untuk melakukan seleksi selektifitas dari senyawa yang diuji. Tahapan kedua ini akan meningkatkan akurasi dengan menggunakan metode *machine learning*. Model ini dibuat dengan metode *Neural network* dengan penambahan *multi-tasking*, *cascaded*, dan *three way* yang dibandingkan dengan *baseline model structure-selectivity-relationship* (SSR).

Penambahan *multi-tasking*, *cascaded*, dan *three way* meningkatkan performa dari model *baseline*, terutama pada skema target protein tunggal. Meskipun begitu, pada skema target lebih dari satu, performanya menurun sehingga diperlukan kajian lebih lanjut (Ning dan Karypis, 2012).

Salah satu tahapan penting dalam kajian pengembangan obat adalah mulai digunakannya metode *deep learning*. Sebagai bagian dari metode *machine learning*, metode ini menggunakan basis *artificial neural network* yang terinspirasi dari simulasi jaringan saraf di otak manusia maupun bagian tubuh lainnya (Schmidhuber, 2015). Aplikasi *deep learning* dalam penapisan senyawa anti-kanker memang belum sebanyak dan sebesar metode *machine learning* lainnya karena kajian ini masih sangat baru, namun sudah mulai ada pengembangan dan penggunaannya. Terdapat beberapa aplikasi *deep learning* di kajian ini. Penggunaan profil obat kanker

sebagai model *deep learning* yang memprediksi efektifitas obat kanker secara genomik merupakan salah satu aplikasi pada ranah *personalized medicine* dengan menggunakan data galur sel manusia dan struktur obat untuk melakukan simulasi molekuler (Chang *et al.*, 2018).

Kemudian, penggunaan *deep learning* untuk mengetahui efek sinergistik dari obat kanker dengan menggunakan data *chemo* dan *genomics informatics* (Preuer *et al.*, 2018). Perbedaan aplikasi *deep learning* dengan metode *machine learning* lainnya seperti *Random Forest* dan *SVM* adalah penggunaan indikator data yang lebih banyak dan lebih kompleks dalam konteks jaringan (*networking*). *Deep learning* lebih cocok diaplikasikan untuk struktur data yang kompleks dan terkoneksi satu sama lain. Hal ini terbukti dengan efektifitas penggunaannya pada *crowd-sourced QSAR* dan prediksi toksikologi terhadap senyawa bioaktif, yang melibatkan banyak nodus di jaringan internet (Koutsoukas *et al.*, 2017).

## KESIMPULAN DAN SARAN

### Kesimpulan

Metode *machine learning* ternyata sangat bermanfaat untuk melakukan penjarangan kepustakaan senyawa bahan alam dan turunannya, baik data berukuran kecil/sedang, maupun berukuran besar (*big data*). Diketahui bahwa metode *machine learning* paling efektif dan efisien adalah *Naive Bayes*. Namun yang terbaik dalam mengolah data senyawa dalam jumlah besar adalah *SVM classifier*. Metode *deep learning* sudah mulai digunakan untuk komputasi data senyawa bioaktif yang kompleks, walaupun aplikasinya masih terbatas.

### Saran

Pengembangan basis data bahan alam dapat diarahkan untuk menjadi repositori data *wet laboratory* bahan alam dengan kapabilitas prediktif berbasis *machine learning*, terutama *deep learning* sehingga bisa menjadi pusat rujukan anotasi untuk eksperimen berikutnya. Jika data yang ada strukturnya sangat kompleks dan membentuk suatu topologi jaringan, maka disarankan menggunakan metode *deep learning* untuk komputasinya. Untuk memperdalam telaah dalam kajian *deep learning*, dapat dilakukan penerapan metode Prisma untuk kedepannya.

## UCAPAN TERIMA KASIH

Terima kasih kami haturkan kepada Ketua LPPM i3L Surjawan. PhD dan Wakil Rektor I i3L Matteo Morello. ,PhD atas bantuan dan dukungannya terhadap kajian ini melalui Hibah Penguatan Prodi S1 Bioinformatika. Terima kasih juga kami haturkan kepada Direktorat Riset dan Pengabdian Masyarakat, Direktorat Jenderal Penguatan Riset dan Pengembangan Kementerian

Riset, Teknologi, dan Pendidikan Tinggi Republik Indonesia dengan bantuan Hibah Penelitian Berbasis Kompetensi DIKTI/KOPERTIS III 2018 No. 049/KM/PNT/2018 dan Hibah Penelitian Dasar DIKTI No.T/140/E3/RA.00/2019.

#### DAFTAR PUSTAKA

- Abraham DJ. 2003. *Burger's Medicinal Chemistry and Drug Discovery, Volume 1: Drug Discovery*, 6<sup>th</sup> ed. New York: Willey Interscience.
- Amin AHM, Khan AI. 2013. One-Shot Classification of 2-D Leaf Shapes using Distributed Hierarchical Graph Neuron (DHGN) Scheme with k-NN Classifier. *Procedia Computer Science*. 24: 84-96.
- Anassamy T, Ahmad BA, Mohd AS, Siddig IA, Syam M, Behnam K. 2013. A Phenylbutenoid Dimer, cis-3- (3', 4'-Dimethoxyphenyl)-4-[(E)-3''', 4'''-Dimethoxystyryl] Cyclohex-1-ene, Exhibits Apoptogenic properties in T-Acute Lymphoblastic Leukemia Cells via induction of p53-Independent Mitochondrial Signalling Pathway. *Evid Based Complement Alternat Med*. 939810.
- Arai K, Abdullah IN, Okumura H. 2013. Identification of Ornamental Plant Functioned as Medicinal Plant Based on Redundant Discrete Wavelet Transformation. *International Journal of Advanced Research in Artificial Intelligence*. 2(3): 60-64.
- Arbab IA, Abdul AB, Sukari MA, Abdullah R, Syam S, et al. 2013. Dentatin isolated from *Clausena excavata* induces apoptosis in MCF-7 cells through the intrinsic pathway with involvement of NFkB signalling and G0/G1 cell cycle arrest: a bioassay-guided approach. *J Ethnopharmacol*. 145: 343-54.
- Backes AR, Casanova D, Bruno OM. 2009. Plant Leaf Identification based on Volumetric Fractal Dimension. *Intern J Pattern Recognit Artif Intell*. 23(6): 145-1160.
- Begue A, Kowlessur V, Singh U, Mahomodally F, Pudaruth S. 2017. Automatic Recognition of Medicinal Plants using Machine Learning Techniques. *IJACSA*. 8(4). DOI : 10.14569/IJACSA.2017.080424.
- Bundela S, Sharma A, Bisen PS. 2015. Potential compounds for oral cancer treatment: Resveratrol, nimbolide, lovastatin, bortezomib, vorinostat, berberine, pterostilbene, deguelin, andrographolide, and colchicine. *PLoS One* [Internet]. Public Library of Science; 10(11):e0141719. Available from: <http://dx.plos.org/10.1371/journal.pone.0141719>
- Carranza-Rojas J, Mata-Montero E. 2016. Combining Leaf Shape and Texture for Costa Rican Plant Species Identification. *CLEIej*. 19(1) Montevideo abr. 2016
- Chaki J. Parekh R, Bhattacharya S. 2015. Plant leaf recognition using texture and shape features with neural classifiers. *Pattern Recognit. Lett*. 58: 61-68.
- Chang Y, Park H, Yang HJ, Lee S, Lee KY, Kim TS, Jung J, Shin JM. 2018. Cancer Drug Response Profile scan (CDRscan): A Deep Learning Model That Predicts Drug Effectiveness from Cancer Genomic Signature. *Scientific Reports*. 8: 8857. DOI:10.1038/s41598-018-27214-6
- Chrystal K, Allan S, Forgeson G, Isaacs R. 2003. The use of complementary/alternative medicine by cancer patients in a New Zealand regional cancer treatment centre. *N Z Med J*. 116(1168):U296.
- Departemen Kesehatan Republik Indonesia. 2007. Kebijakan Obat Tradisional Nasional Tahun 2007. Departemen Kesehatan Republik Indonesia. [http://binfar.depkes.go.id/dat/lama/1206328790\\_Buku%20Kebijakan%20Obat%20Tradisional%20Nasional%20Tahun%202007.pdf](http://binfar.depkes.go.id/dat/lama/1206328790_Buku%20Kebijakan%20Obat%20Tradisional%20Nasional%20Tahun%202007.pdf).
- Ding L, Wang ZZ, Sun XD, Yang J, Ma CY, Li W, Liu HM. 2017. 3D-QSAR (CoMFA, CoMSIA), molecular docking and molecular dynamics simulations study of 6-aryl-5-cyanopyrimidine derivatives to explore the structure requirements of LSD1 inhibitors. *Bioorg Med Chem Lett*. 27(15): 3521-8. doi: 10.1016/j.bmcl.2017.05.065.
- Du JX, Zhai CM, Wang QP. 2013. Recognition of plant leaf image based on fractal dimension features. *Neurocomputing*. 116: 150-156. <https://doi.org/10.1016/j.neucom.2012.03.028>
- Du M, Zhang S, dan Wang H. 2009. Supervised Isomap for Plant Leaf Image Classification. *5th International Conference on Emerging Intelligent Computing Technology and Applications*. Ulsan, South Korea, 16-19 September 2009.
- Foo JB, Yazan LS, Chan KW, Tahir PM, Ismail M. 2011. Kenaf seed oil from supercritical carbon dioxide fluid extraction induced G1 phase cell cycle arrest and apoptosis in leukemia cells. *Afr. J. Biotechnol*. 10: 5389-5397
- Golbeck J, Frago G, Hartel F, Hendler J, Oberthaler J, Parsia B. 2011. The national Cancer institute's thesaurus and ontology. *Web semantic: Science, Services and Agents on the World Wide Web*. 1(1).
- Han C, Li Z, Hou J, Wang Z, Xu D, Xue G, Kong, L. 2018. Bioactivity evaluation of natural product  $\alpha$ -mangostin as a novel xanthone-based lysine-specific demethylase 1 inhibitor to against tumor metastasis. *Bioorg. Chem.*, 76, 415-419. <https://doi.org/10.1016/J.BIOORG.2017.12.004>



- Hawkins P dan Skillman G. 2006. Ligand-based design workflow. [http://images.apple.com/science/pdf/ligandbased\\_design\\_workflow.pdf](http://images.apple.com/science/pdf/ligandbased_design_workflow.pdf).
- Herdiyeni Y dan Wahyuni NKS. 2012. Mobile Application for Indonesian Medicinal Plants Identification using Fuzzy Local Binary Pattern and Fuzzy Color Histogram. *International Conference on Advanced Computer Science and Information Systems (ICACSIS)*. West Java, Indonesia. 301-306 December 2012.
- Hernandez-Serna A dan Jimenez-Segura LF. 2014. Automatic Identification of species with neural networks. *Peer J* 2:e563. doi:10.7717/peerj.563.
- Ho WY, Yeap SK, Ho CL, Raha AR, Suraini AA2011. *Elephantopus scaber* induces cytotoxicity in MCF-7 human breast cancer cells via p53-induced apoptosis. *J. Med. Plants Res.* 5: 5741-5749.
- Ibrahim MY, Hashim NM, dan Mohan S. 2015.  $\alpha$ -Mangostin from *Cratoxylum arborescens*: an in vitro and in vivo toxicological evaluation. *Arab. J. Chem.* 8: 129–137.
- Isa NM, Abdul AB, Abdelwahab SI, Abdullah R, Sukari MA. 2012. Boesenbergin A, a chalcone from *Boesenbergia rotunda* induces apoptosis via mitochondrial dysregulation and cytochrome c release in A549 cells in vitro: Involvement of HSP70 and Bcl2/Bax signalling pathways. *J Funct Foods.* 5: 87-97.
- Jamal S dan Scaria V. 2013. Cheminformatic models based on machine learning for pyruvate kinase inhibitors of *Leishmania mexicana*. *BMC Bioinformatics.* *BioMed Central.* 14(1): 329. <https://doi.org/10.1186/1471-2105-14-329>.
- Jayaram B. 2011. SCFBIO: What is drug design? <http://www.scfbio-iitd.res.in/tutorial/drugdiscovery.htm>.
- Kim K, Lee JM, Yu YS, Kim H, Nam HJ, Moon HG, et al. 2017. ROR $\alpha$ 2 requires LSD1 to enhance tumor progression in breast cancer. *Sci Rep.* 7(1): 11994. DOI:10.1038/s41598-017-12344-0.
- Koutsoukas A, Monaghan KJ, Li X, Huan J. 2017. Deep-learning: Investigating deep neural networks hyper-parameters and comparison of performance to shallow methods for modeling bioactivity data. *J Cheminform.* 9(1): 42. <https://doi.org/10.1186/s13321-017-0226-y>.
- Kukreja M, Johnston SA, dan Stafford P. 2012. Comparative study of classification algorithms for immunosignaturing data. *BMC Bioinformatics.* 13:139. <https://doi.org/10.1186/1471-2105-13-139>.
- Lampe JW. 1999. Health effects of vegetables and fruit: assessing mechanisms of action in human experimental studies. *Am J Clin Nutr.* 70: 475s-490s.
- Le TL, Tran DT, Hoang VN. 2014. Fully Automatic leaf-based plant identification, application for Vietnamese medicinal plant search. *Fifth Symposium on Information and Communication Technology*. Hanoi, Vietnam, 146-154.
- Lim GCC. 2002. Overview of cancer in Malaysia. *Jpn. J. Clin. Oncol.* 32: S37-S42.
- Lim S, Janzer A, Becker A. 2010. Lysinespecific demethylase 1 (LSD1) is highly expressed in ER-negative breast cancers and a biomarker predicting aggressive biology. *Carcinogenesis.* 31: 512–520
- Lynch JT, Cockerill MJ, dan Hitchin JR. 2013. CD86 expression as a surrogate cellular biomarker for pharmacological inhibition of the histone demethylase lysine-specific demethylase 1. *Anal. Biochem.* 442: 104–106
- Mohammad H, Smitheman K, dan Cusan M. 2013. Inhibition of LSD1 as a therapeutic strategy for the treatment of acute myeloid leukemia. *Blood.* 122: 3964.
- Mohan S, Abdul AB, Abdelwahab SI, Al-Zubairi AS, Aspollah SM, et al. 2010a Typhonium flagelliforme inhibits the proliferation of murine leukemia WEHI-3 cells in vitro and induces apoptosis in vivo. *Leuk Res.* 34: 1483-1492.
- Ma QS, Yao Y, Zheng YC, Feng S, Chang J, Yu B, et al. 2019. Ligand-based design, synthesis and biological evaluation of xanthine derivatives as LSD1/KDM1A inhibitors. *Eur J Med Chem.* 162: 555–67. <https://doi.org/10.1016/j.ejmech.2018.11.035>.
- Mohan S, Bustamam AA, Ibrahim S, Al-Zubairi AS, Aspollah M, et al. 2010b. In Vitro Ultramorphological Assessment of Apoptosis on CEMss Induced by Linoleic Acid-Rich fraction from Typhonium flagelliforme Tuber. *Evid Based Complement Alternative Med.* 421894-421894
- Munisami T, Ramsurn M, Kishnah S, Pudaruth S. 2015. Plant leaf recognition using shape features and colour histogram with k-nearest neighbour classifiers. *Procedia Computer Science.* 58: 740-747.
- Murphy RF. 2011. An active role for machine learning in drug development. *Nature Chemical Biology.* 7: 327-330.
- Naik AW, Kangas JD, Langmead CJ, Murphy RF. 2013. Efficient Modeling and Active Learning Discovery of Biological Responses. *PLoS ONE.* 8: e83996.
- Nand M, Maiti P, Pant R, Kumari M, Chandra S, Pande V. 2016. Virtual screening of natural compounds as inhibitors of EGFR 696-1022 T790M associated with non-small cell lung cancer. *Bioinformation.* 12(06): 311–317. doi: 10.6026/97320630012311.

- Negi JS, Bisht VK, dan Singh P. 2013. Naturally occurring xanthenes: chemistry and biology. *Journal Applied Chemical*. 1–9. <http://dx.doi.org/10.1155/2013/621459>.
- Ngo ST dan Li MS. 2013. Top-leads from natural products for treatment of Alzheimer's disease: docking and molecular dynamics study. *Molecular Simulation*. 39 (4): 279–291.
- Ning X dan Karypis G. 2012. Improved Machine Learning Models for Predicting Selective Compounds. *Journal Chem Inf Model*. 52(1):38–50. doi: 10.1021/ci200346b.
- Parikesit AA. 2018. Kontribusi Aplikasi Medis dari Ilmu Bioinformatika Berdasarkan Perkembangan Pembelajaran Mesin (Machine Learning) Terbaru. *Cermin Dunia Kedokt*; 45(9): 700–3. Available from: <http://www.kalbemed.com/Desktop Modules/EasyDNNNews/DocumentDownload.ashx?portalid=0&moduleid=471&articleid=225&documentid=65>
- Parikesit AA, Nurdiansyah R, dan Agustriawan D. 2018. Telaah Sistematis Terhadap Basis Data Bahan Alam untuk Pengembangan Produk Suplemen Herbal. *Pros SEMNASTAN*; 0(0):62–8. Available from: <https://jurnal.umj.ac.id/index.php/semnastan/article/view/2259/1874>
- Prasvita DS dan Herdiyeni Y. 2013. MedLeaf: Mobile Application for Medicinal Plant Identification Based on Leaf Image. *IJASEIT*. 3: 5–8.
- Preuer K, Lewis RPI, Hochreiter S, Bender A, Bulusu KC, Klambauer G. 2018. DeepSynergy: Predicting anti-cancer drug synergy with Deep Learning. *Bioinformatics*. 34(9): 1538–1546. DOI: 10.1093/bioinformatics/btx806.
- Ramanto KN dan Parikesit AA. 2019. The usage of deep learning algorithm in medical diagnostic of breast cancer. *Malaysian Journal Fundam Appl Sci*. 15(2): 274–281.
- Reker D, Rodrigues T, Schneider P, Schneider G. 2014. Identifying the macromolecular targets of de novo-designed chemical entities through self-organizing map consensus. *Proc Natl Acad Sci*. 111(11): 4067–72. <https://doi.org/10.1073/pnas.1320001111>.
- Schmitt ML, Hauser AT, dan Carlino L. 2013. Nonpeptidic propargylamines as inhibitors of lysine specific demethylase 1 (LSD1) with cellular activity. *Journal Med. Chem*. 56: 7334–7342
- Schmidhuber J. 2015. Deep learning in neural networks: An overview. *Neural Networks. Pergamon*. 61:85–117. Available from: <https://www.sciencedirect.com/science/article/pii/S0893608014002135?via%3Dihub>.
- Schneider G, Reker D, Chen T, Hauenstein K, Schneider P, Altmann K-H. 2016. Deorphaning the Macromolecular Targets of the Natural Anticancer Compound Dolicolide. *Angew Chemie Int Ed*. 55(40): 12408–11. <http://doi.wiley.com/10.1002/anie.201605707>
- Schulte JH, Lim S, Schramm A. 2009. Lysine-specific demethylase 1 is strongly expressed in poorly differentiated neuroblastoma: implications for therapy. *Cancer Res*. 69: 2065–2071.
- Siravenha, ACQ, Carvalho SR. 2015. Exploring the use of Leaf Shape Frequencies for Plant Classification. *28th SIBGRAPI Conference on Graphics, Patterns and Images*, Salvador, Brazil, 297–304.
- Surh YJ. 2003. Cancer chemoprevention with dietary phytochemicals. *Nat Rev Cancer*. 3: 768–780.
- Syam S, Abdul AB, Sukari MA, Mohan S, Abdelwahab SI, et al. 2011. The growth suppressing effects of girinimbine on HepG2 involve induction of apoptosis and cell cycle arrest. *Molecules*. 16: 7155–7170.
- Wang Y, Zhang H, Chen Y. 2009. LSD1 is a subunit of the NuRD complex and targets the metastasis programs in breast cancer. *Cell*. 138: 660–672.
- Wissmann M, Yin N, Muller JM. 2007. Cooperative demethylation by JMJD2C and LSD1 promotes androgen receptor-dependent gene expression. *Nat. Cell Biol*. 9: 347–353.
- Wahi D, Jamal S, Goyal S, Singh A, Jain R, Rana P, et al. 2015. Cheminformatics models based on machine learning approaches for design of USP1/UAF1 abrogators as anticancer agents. *Syst Synth Biol*. 9(1–2): 33–43. <http://link.springer.com/10.1007/s11693-015-9162-1>
- Wu SG, Bao FS, Xu EY, Wang Y, Chang YF, Xiang QL. 2007. A Leaf Recognition Algorithm for Plant Classification Using Probabilistic Neural Network. *IEEE 7th International Symposium on Signal Processing and Information Technology*. Cairo, Egypt. Dec. 2007.
- Yanuar, A., Mun'im, A., Lagho, A. B. A., Syahdi, R. R., Rahmat, M., & Suhartanto, H. 2011. Medicinal Plants Database and Three Dimensional Structure of the Chemical Compounds from Medicinal Plants in Indonesia. *International Journal of Computer Science*. 8(5), 180–183. Biomolecules. <http://arxiv.org/abs/1111.7183>
- Yu S, Liu Y, Wang G, Zhang H. 2017. Deep Learning for Plant Identification in Natural Environment. *Hindawi Computational Intelligence and Neuroscience*. Article ID 7361042. <https://doi.org/10.1155/2017/7361042>.